



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2020 – 08:04 pm BST

PDB ID : 3AVN  
Title : Crystal structures of novel allosteric peptide inhibitors of HIV integrase in the LEDGF binding site  
Authors : Peat, T.S.; Deadman, J.J.; Newman, J.; Rhodes, D.I.  
Deposited on : 2011-03-05  
Resolution : 2.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

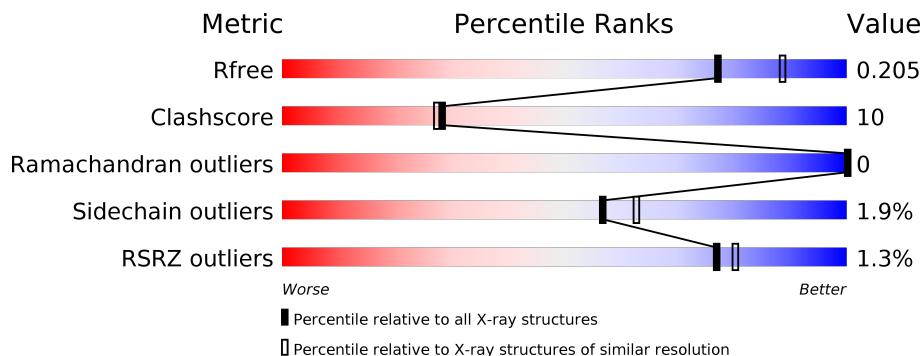
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	183	 13% 69% 12% 19%
1	B	183	 66% 15% 19%
2	G	8	 13% 75% 13% 13%
2	H	8	 13% 75% 13% 13%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	ACY	B	213	-	-	X	-

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 2742 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Integrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	149	1216	779	207	225	5	0	10	0
1	B	149	1241	797	211	228	5	0	13	0

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	MET	-	EXPRESSION TAG	UNP P12497
A	31	GLY	-	EXPRESSION TAG	UNP P12497
A	32	SER	-	EXPRESSION TAG	UNP P12497
A	33	SER	-	EXPRESSION TAG	UNP P12497
A	34	HIS	-	EXPRESSION TAG	UNP P12497
A	35	HIS	-	EXPRESSION TAG	UNP P12497
A	36	HIS	-	EXPRESSION TAG	UNP P12497
A	37	HIS	-	EXPRESSION TAG	UNP P12497
A	38	HIS	-	EXPRESSION TAG	UNP P12497
A	39	HIS	-	EXPRESSION TAG	UNP P12497
A	40	SER	-	EXPRESSION TAG	UNP P12497
A	41	SER	-	EXPRESSION TAG	UNP P12497
A	42	GLY	-	EXPRESSION TAG	UNP P12497
A	43	LEU	-	EXPRESSION TAG	UNP P12497
A	44	VAL	-	EXPRESSION TAG	UNP P12497
A	45	PRO	-	EXPRESSION TAG	UNP P12497
A	46	ARG	-	EXPRESSION TAG	UNP P12497
A	47	GLY	-	EXPRESSION TAG	UNP P12497
A	48	SER	-	EXPRESSION TAG	UNP P12497
A	49	HIS	-	EXPRESSION TAG	UNP P12497
A	56	SER	CYS	ENGINEERED MUTATION	UNP P12497
A	139	ASP	PHE	ENGINEERED MUTATION	UNP P12497
A	185	HIS	PHE	ENGINEERED MUTATION	UNP P12497
B	30	MET	-	EXPRESSION TAG	UNP P12497
B	31	GLY	-	EXPRESSION TAG	UNP P12497

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	32	SER	-	EXPRESSION TAG	UNP P12497
B	33	SER	-	EXPRESSION TAG	UNP P12497
B	34	HIS	-	EXPRESSION TAG	UNP P12497
B	35	HIS	-	EXPRESSION TAG	UNP P12497
B	36	HIS	-	EXPRESSION TAG	UNP P12497
B	37	HIS	-	EXPRESSION TAG	UNP P12497
B	38	HIS	-	EXPRESSION TAG	UNP P12497
B	39	HIS	-	EXPRESSION TAG	UNP P12497
B	40	SER	-	EXPRESSION TAG	UNP P12497
B	41	SER	-	EXPRESSION TAG	UNP P12497
B	42	GLY	-	EXPRESSION TAG	UNP P12497
B	43	LEU	-	EXPRESSION TAG	UNP P12497
B	44	VAL	-	EXPRESSION TAG	UNP P12497
B	45	PRO	-	EXPRESSION TAG	UNP P12497
B	46	ARG	-	EXPRESSION TAG	UNP P12497
B	47	GLY	-	EXPRESSION TAG	UNP P12497
B	48	SER	-	EXPRESSION TAG	UNP P12497
B	49	HIS	-	EXPRESSION TAG	UNP P12497
B	56	SER	CYS	ENGINEERED MUTATION	UNP P12497
B	139	ASP	PHE	ENGINEERED MUTATION	UNP P12497
B	185	HIS	PHE	ENGINEERED MUTATION	UNP P12497

- Molecule 2 is a protein called LEDGF peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	G	8	Total	C	N	O	0	0	0
			65	39	12	14			
2	H	8	Total	C	N	O	0	0	0
			65	39	12	14			

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).

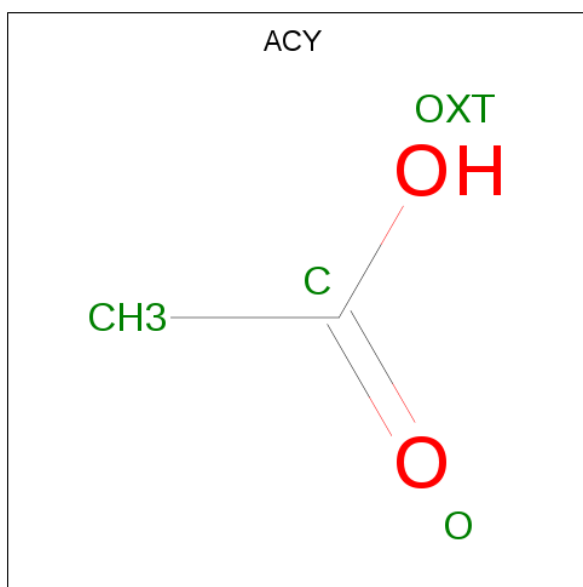


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Cl 1 1	0	0
4	A	1	Total Cl 1 1	0	0

- Molecule 5 is ACETIC ACID (three-letter code: ACY) (formula: C<sub>2</sub>H<sub>4</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C O 4 2 2	0	0
5	B	1	Total C O 4 2 2	0	0

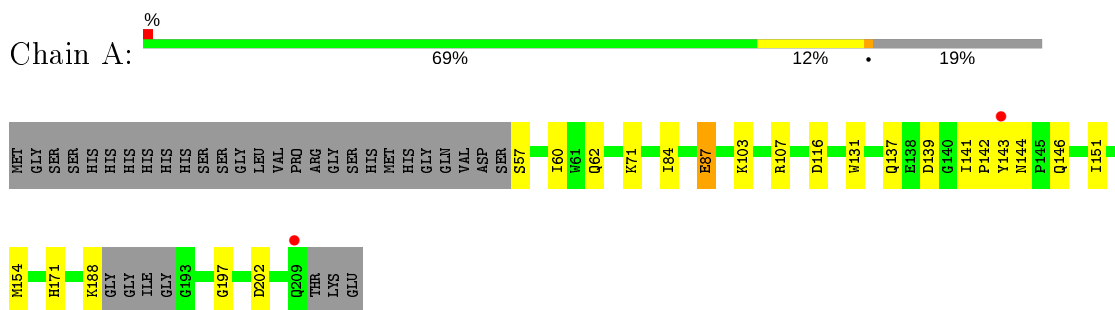
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	48	Total O 48 48	0	0
6	B	53	Total O 53 53	0	0
6	G	2	Total O 2 2	0	0
6	H	2	Total O 2 2	0	0

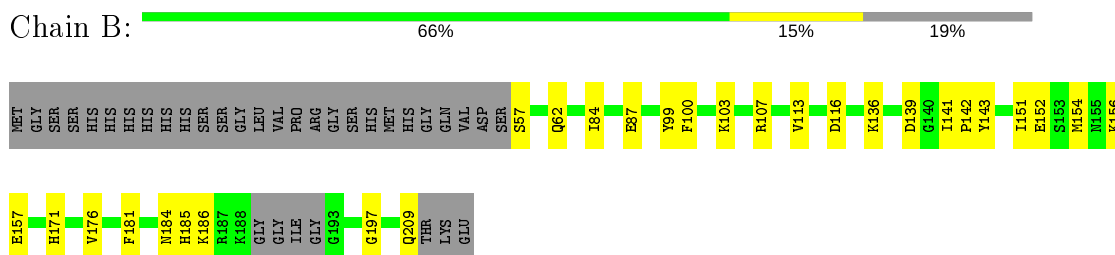
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

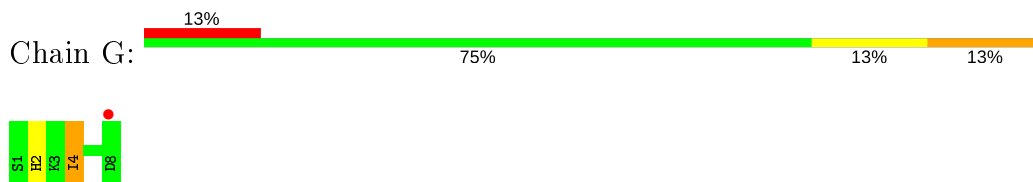
- Molecule 1: Integrase



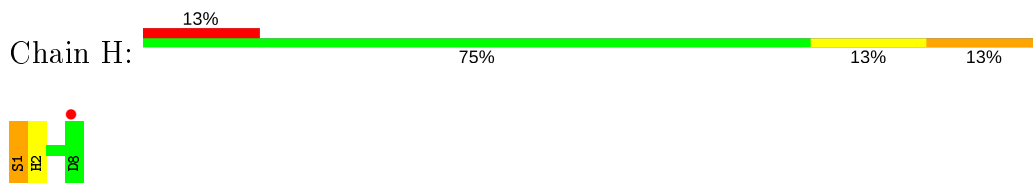
- Molecule 1: Integrase



- Molecule 2: LEDGF peptide



- Molecule 2: LEDGF peptide





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.57Å 70.57Å 67.47Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	61.11 – 2.10 45.30 – 2.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (61.11-2.10) 100.0 (45.30-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.03 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.6.0062	Depositor
R, $R_{free}$	0.158 , 0.205 0.157 , 0.205	Depositor DCC
$R_{free}$ test set	1125 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.7	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 34.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l 0.479 for h,-h-k,-l 0.025 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2742	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.96% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ACY, SO4, CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	1.26	4/1264 (0.3%)	0.93	2/1713 (0.1%)
1	B	1.23	4/1289 (0.3%)	0.92	1/1746 (0.1%)
2	G	0.77	0/65	1.06	1/86 (1.2%)
2	H	0.78	0/65	0.94	0/86
All	All	1.23	8/2683 (0.3%)	0.93	4/3631 (0.1%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	107	ARG	CZ-NH1	5.52	1.40	1.33
1	B	157	GLU	CG-CD	5.44	1.60	1.51
1	B	100	PHE	CD1-CE1	5.30	1.49	1.39
1	A	131	TRP	CB-CG	5.29	1.59	1.50
1	A	87[A]	GLU	CG-CD	5.11	1.59	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	116	ASP	CB-CG-OD1	7.38	124.94	118.30
1	A	116	ASP	CB-CG-OD1	6.11	123.80	118.30
1	A	107	ARG	NE-CZ-NH2	-5.95	117.33	120.30
2	G	4	ILE	CG1-CB-CG2	-5.22	99.91	111.40

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1216	0	1242	25	0
1	B	1241	0	1275	31	0
2	G	65	0	61	1	0
2	H	65	0	61	3	0
3	A	20	0	0	1	0
3	B	20	0	0	1	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	B	8	0	6	4	0
6	A	48	0	0	3	0
6	B	53	0	0	3	0
6	G	2	0	0	1	0
6	H	2	0	0	0	0
All	All	2742	0	2645	52	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

The worst 5 of 52 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:87[B]:GLU:OE1	6:B:241:HOH:O	1.80	0.98
1:A:87[B]:GLU:OE1	6:A:238:HOH:O	1.81	0.96
1:A:202[B]:ASP:OD1	1:B:209:GLN:NE2	2.08	0.87
1:A:141[B]:ILE:HD12	1:A:142:PRO:HD2	1.66	0.78
1:B:141[B]:ILE:HD12	1:B:142:PRO:HD2	1.67	0.76

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries

of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	155/183 (85%)	154 (99%)	1 (1%)	0	100	100
1	B	158/183 (86%)	157 (99%)	1 (1%)	0	100	100
2	G	6/8 (75%)	5 (83%)	1 (17%)	0	100	100
2	H	6/8 (75%)	4 (67%)	2 (33%)	0	100	100
All	All	325/382 (85%)	320 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	133/150 (89%)	131 (98%)	2 (2%)	65	71
1	B	136/150 (91%)	135 (99%)	1 (1%)	84	88
2	G	8/8 (100%)	7 (88%)	1 (12%)	4	2
2	H	8/8 (100%)	7 (88%)	1 (12%)	4	2
All	All	285/316 (90%)	280 (98%)	5 (2%)	57	65

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	137	GLN
1	A	188	LYS
1	B	186	LYS
2	G	4	ILE
2	H	1	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	171	HIS
2	H	2	HIS
1	B	185	HIS
1	A	171	HIS
1	B	209	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

Of 12 ligands modelled in this entry, 2 are monoatomic - leaving 10 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	SO4	B	3	-	4,4,4	0.19	0	6,6,6	0.51	0
3	SO4	A	4	-	4,4,4	0.19	0	6,6,6	0.65	0
5	ACY	B	214	-	1,3,3	0.13	0	0,3,3	0.00	-
3	SO4	B	6	-	4,4,4	0.30	0	6,6,6	1.68	1 (16%)
3	SO4	A	7	-	4,4,4	0.70	0	6,6,6	0.47	0
5	ACY	B	213	-	1,3,3	1.17	0	0,3,3	0.00	-
3	SO4	A	5	-	4,4,4	0.72	0	6,6,6	1.53	1 (16%)
3	SO4	B	1	-	4,4,4	0.10	0	6,6,6	0.52	0
3	SO4	A	2	-	4,4,4	0.13	0	6,6,6	0.85	0
3	SO4	B	8	-	4,4,4	0.72	0	6,6,6	0.39	0

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	B	6	SO4	O4-S-O3	3.46	123.84	109.06
3	A	5	SO4	O2-S-O1	2.31	126.47	109.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	3	SO4	1	0
3	A	4	SO4	1	0
5	B	214	ACY	1	0
5	B	213	ACY	3	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	149/183 (81%)	-0.54	2 (1%) 77 80	15, 23, 45, 63	19 (12%)
1	B	149/183 (81%)	-0.59	0 100 100	15, 23, 43, 51	17 (11%)
2	G	8/8 (100%)	0.35	1 (12%) 3 5	33, 40, 66, 67	1 (12%)
2	H	8/8 (100%)	0.44	1 (12%) 3 5	34, 43, 64, 64	1 (12%)
All	All	314/382 (82%)	-0.51	4 (1%) 77 80	15, 23, 48, 67	38 (12%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	8	ASP	2.3
1	A	209	GLN	2.2
1	A	143	TYR	2.1
2	G	8	ASP	2.1

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	SO4	B	8	5/5	0.91	0.16	41,42,52,52	5
3	SO4	B	6	5/5	0.92	0.14	26,30,38,41	5
3	SO4	A	2	5/5	0.93	0.12	35,45,50,61	5
3	SO4	A	7	5/5	0.93	0.13	40,41,52,53	5
5	ACY	B	214	4/4	0.94	0.28	37,43,46,53	0
3	SO4	A	5	5/5	0.94	0.13	23,28,36,38	5
5	ACY	B	213	4/4	0.96	0.21	33,38,44,46	0
4	CL	A	1	1/1	0.97	0.09	55,55,55,55	0
3	SO4	B	1	5/5	0.97	0.10	37,47,55,59	5
4	CL	B	2	1/1	0.97	0.06	57,57,57,57	0
3	SO4	B	3	5/5	0.97	0.09	31,37,44,45	5
3	SO4	A	4	5/5	0.99	0.11	29,30,36,39	5

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.